

Measurement of the Top Quark Mass in the L+jets Channel with the MTM3 Method — Preblessing

Paul Lujan, John Freeman, Angela Galtieri, Jeremy Lys, Jason Nielsen, Igor Volobouev

March 27, 2008

- Method Overview
- Phase Space Integration by Quasi-Monte Carlo
- Transfer Functions, Normalization, and Acceptance
- Event Selection and Samples
- Calibration by Pseudo-Experiments
- Systematic Errors
- Results

MTM Evolution

- Original MTM top quark mass measurement started as a multivariate template method with fancy weighting of jet permutations and JES parameter in the kinematic fit. CDF note 7102 (2004).
- MTM2: Started using non-Gaussian transfer functions. Phase space integration (Matrix Element technique) is used instead of the kinematic fit. “Effective propagators” are introduced to compensate for unrealistic assumptions about jet angular resolution and hadronization. A multivariate m_t -uncorrelated discriminant is used to suppress background. CDF notes 8173 (2006), 8780.
- MTM2.5: Neural net discriminant is introduced. CDF notes 8969 (2007), 9025, 9186, 9196.
- MTM3 (this one): Reduced the number of kinematic assumptions. Phase space integration is performed by Quasi-Monte Carlo. Introduced transfer functions for angular jet resolutions. Proto-jet masses are used as transfer function predictors. P_T -dependent JES parameter. CDF note 9245 (March 2008).

Kinematic Assumptions

- 2 initial state particles and 6 final state particles in the $q\bar{q} \rightarrow t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow \ell\nu + 4$ jets process need 32 variables for complete description of all 4-vectors.
- Masses of the initial state partons and of the final state leptons ≈ 0 . This takes away 4 degrees of freedom.
- Energy-momentum conservation takes away 4 more.
- The momentum of the charged lepton is assumed to be perfectly measured (subtract 3 degrees of freedom).
- The transverse momenta of the initial state partons are assumed to be small in comparison with their longitudinal momenta. This allows for determination of the longitudinal momenta without taking into account their transverse motion. Effectively, this assumption reduces the number of degrees of freedom by 2 because we are still interested in the $t\bar{t}$ transverse momentum but not in the transverse momenta of individual initial state partons.
- $32 - 4 - 4 - 3 - 2 = 19$ degrees of freedom

Signal Likelihood

Likelihood for each event:

$$L(\vec{y} \mid m_t, \Delta_{\text{JES}}) = \frac{1}{N(m_t)} \frac{1}{A(m_t, \Delta_{\text{JES}})} \sum_{i=1}^{24} w_i L_i(\vec{y} \mid m_t, \Delta_{\text{JES}})$$

- \vec{y} are the momenta measured in the detector
- m_t is the pole mass of the top quark
- Δ_{JES} is the JES in the units of standard JES systematic error
- $N(m_t)A(m_t, \Delta_{\text{JES}})$ is the likelihood normalization factor: tree-level cross section times acceptance
- w_i are the permutation weights calculated using tagging probabilities
- L_i are likelihoods for individual permutations

Signal Likelihood (Cont'd)

Likelihood for each permutation:

$$L_i(\vec{y} \mid m_t, \Delta_{\text{JES}}) = \int \frac{f(z_1)f(z_2)}{FF} \text{TF}(\vec{y} \mid \vec{x}, \Delta_{\text{JES}}) |M(m_t, \vec{x})|^2 d\Phi(\vec{x})$$

- \vec{x} are the parton-level momenta
- $f(z)$ are the quark/gluon distribution functions, and FF is the flux factor
- $\text{TF}(\vec{y} \mid \vec{x}, \Delta_{\text{JES}})$ are the transfer functions connecting the parton-level and the detector-level quantities
- M is the matrix element of the process. We use the Kleiss-Stirling matrix element, which includes both $q\bar{q}$ and gg , as well as full spin correlations.
- Φ is the parton-level phase space being integrated over

Phase Space Integration

- In practice, evaluation of a large number of 19-dimensional integrals is difficult (**slow!**). MTM2, MEAT, *etc.* introduced additional kinematic assumptions to reduce the dimensionality.
- Choice of the phase space sampling scheme is crucial for convergence. We use the following integration variables:

M_1^2 , the top mass squared on the hadronic side of the event

M_2^2 , the top mass squared on the leptonic side

m_1^2 , the W mass squared on the hadronic side

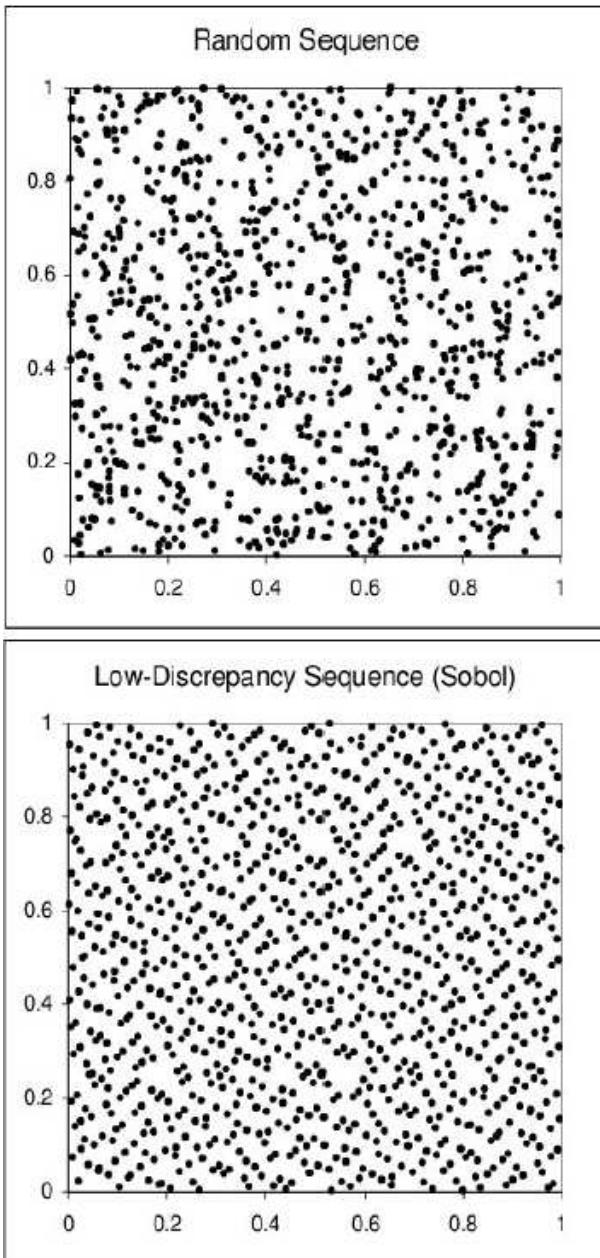
m_2^2 , the W mass squared on the leptonic side

$\beta = \log \frac{\rho_q}{\rho_{\bar{q}}}$, the logarithm of the ratio of the 3-momentum magnitudes of the two partons from the hadronically decaying W

$\bar{p}_T(t\bar{t})$, the transverse momentum of the $t\bar{t}$ system

η , ϕ , and m of each proto-jet

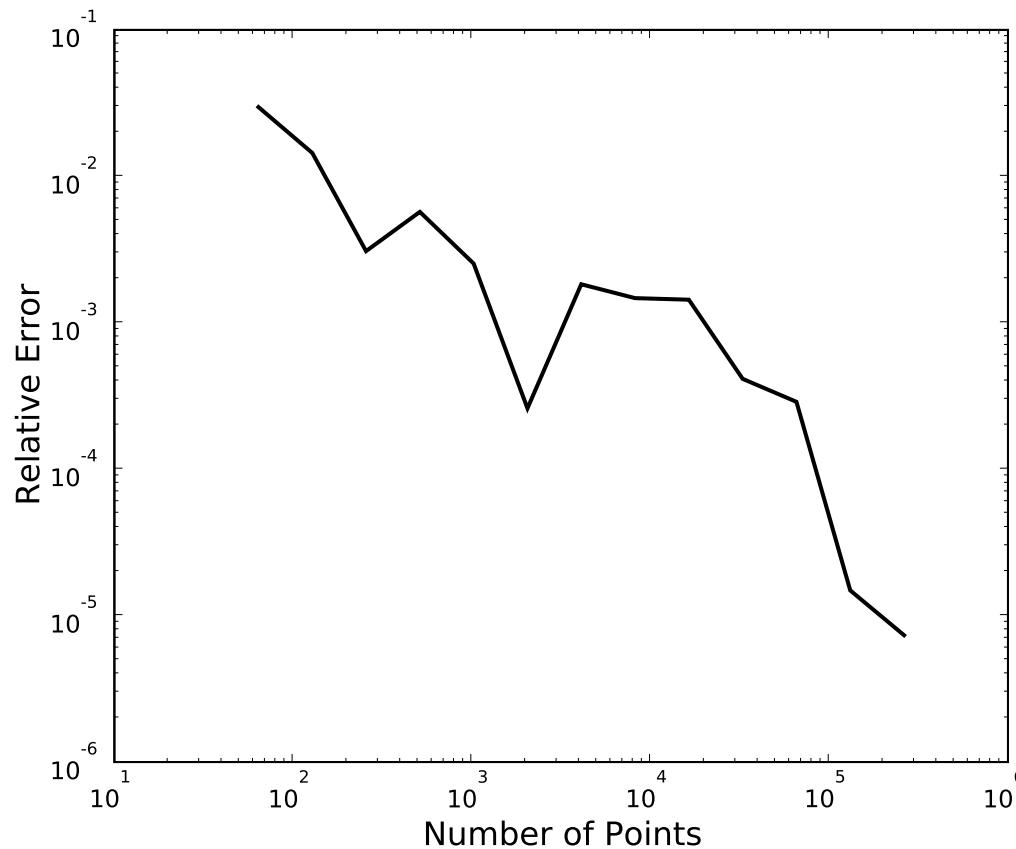
Quasi-Monte Carlo (QMC)



- Best integration method in high dimensions. Started seeing significant practical use in late 80s. To our knowledge, the first study of QMC for HEP-related integrals was published in 2006 (by Kleiss and Lazopoulos).
- Quasi-MC integration uses “low-discrepancy” sequences (we use a variant of the Sobol sequence, plotted on the left) to provide more uniform coverage of the phase space.
- For “well-behaved” functions, convergence rate is guaranteed to be at least as good as $O(\log(N)^d/N)$. Compare with $O(1/\sqrt{N})$ for standard MC.
- We use QMC for 18 dimensions out of 19. Convergence is estimated empirically, from the smoothness of the likelihood curves.

QMC Convergence for a Simple Example

- Muon lifetime calculated using QMC integration over the 5-d phase space of the final state particles. Compared with the exact theoretical result. Apparent convergence rate is close to $O(N^{-1})$.



Speeding Up the Integration

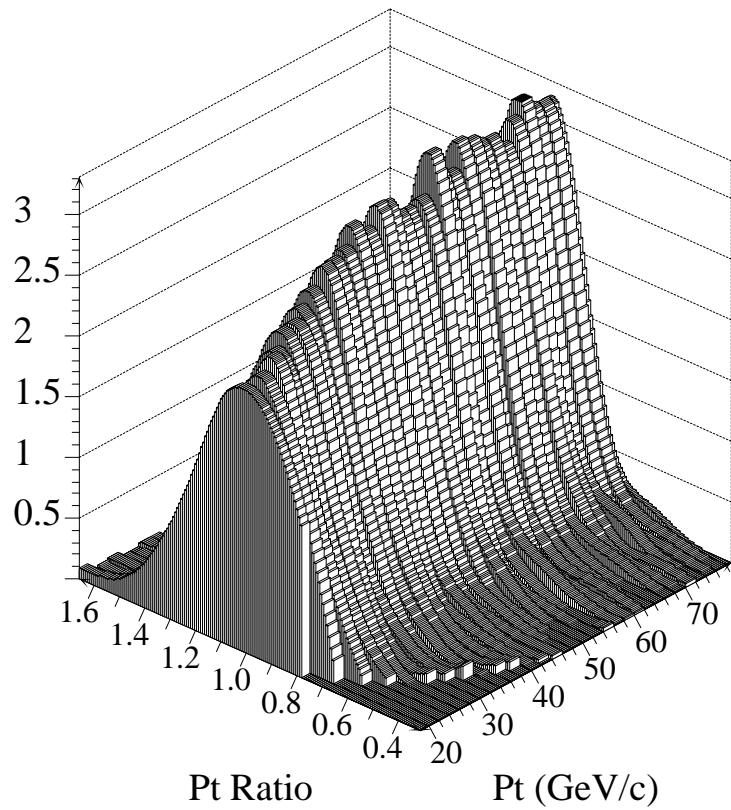
We employ a variety of techniques to further speed up the integration:

- Importance sampling is used in almost every variable. For example, the top masses and the hadronic W mass are sampled in the space of cumulative distributions of their respective Breit-Wigners.
- Smaller top mass scan range (32 2-GeV steps from 143 to 207 GeV) and larger JES scan steps (25 0.4 σ steps from -5σ to $+5\sigma$), 1/2 the size of the MTM2 scans in each variable.
- Quickly identify permutations with lower likelihood and spend less time integrating these permutations.
- The integration terminates when either a preset convergence target or a timeout (currently 2 hours) is reached. About 2/3 of the events reach their target before timing out.
- Overall, the average integration time is about 80 min/event — long but doable.

Transfer Functions

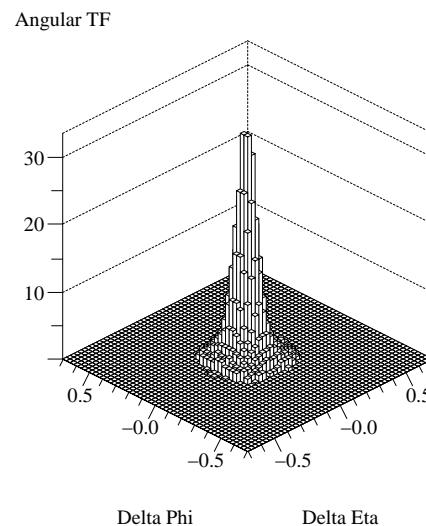
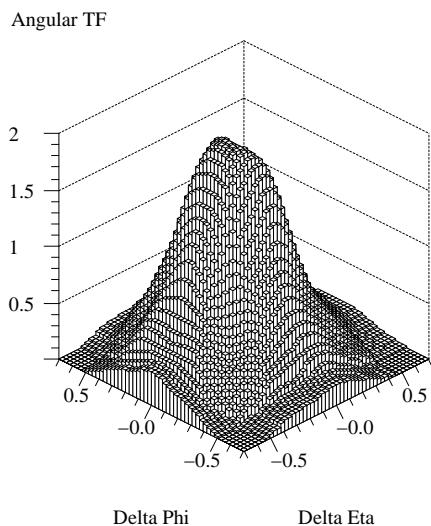
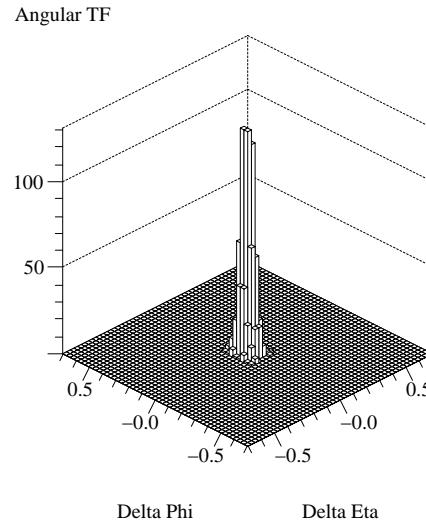
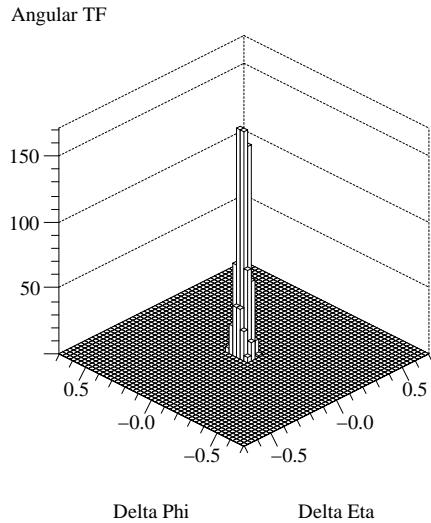
- Transfer functions are factorized into the product of P_T and angular terms

Transfer Function



- P_T TFs are built as a function of the jet-to-parton P_T ratio in different bins of parton P_T , η , and proto-jet mass. Density reconstruction is performed by using local orthogonal polynomial expansion (a kind of a marriage between KDE and orthogonal series estimator).
- The transfer functions are no longer parametrized — they are calculated using an interpolated lookup table (large but reasonably fast).

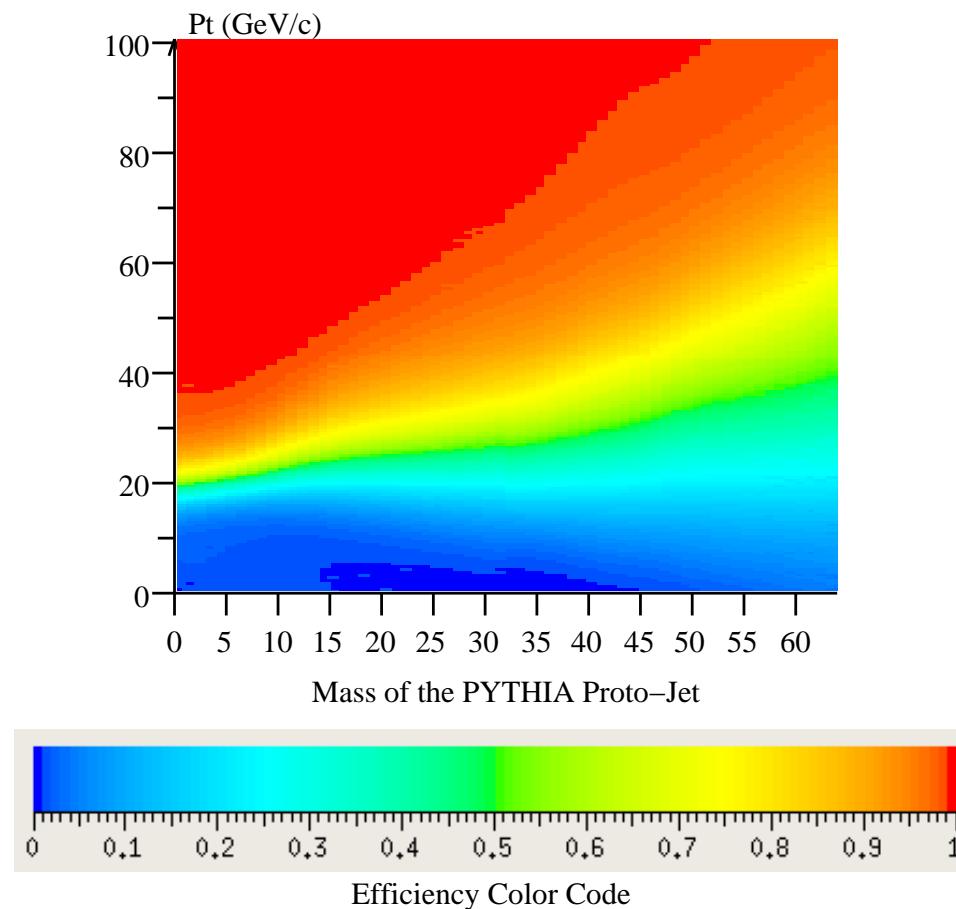
Angular Transfer Functions



- Top left: $M_{pj} = 5 \text{ GeV}/c^2$, $p_{T\text{parton}} = 40 \text{ GeV}/c$, q jet.
- Top right: $M_{pj} = 5 \text{ GeV}/c^2$, $p_{T\text{parton}} = 40 \text{ GeV}/c$, b jet.
- Bottom left: $M_{pj} = 30 \text{ GeV}/c^2$, $p_{T\text{parton}} = 40 \text{ GeV}/c$, q jet.
- Bottom right: $M_{pj} = 30 \text{ GeV}/c^2$, $p_{T\text{parton}} = 100 \text{ GeV}/c$, q jet.

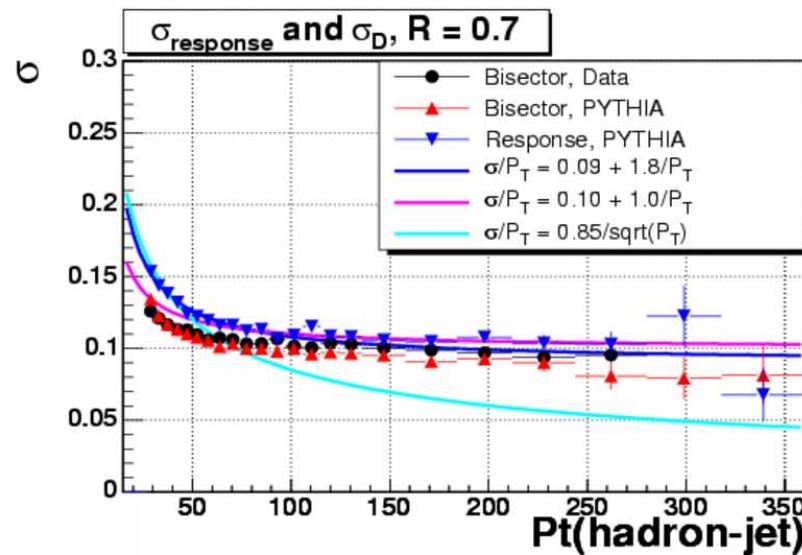
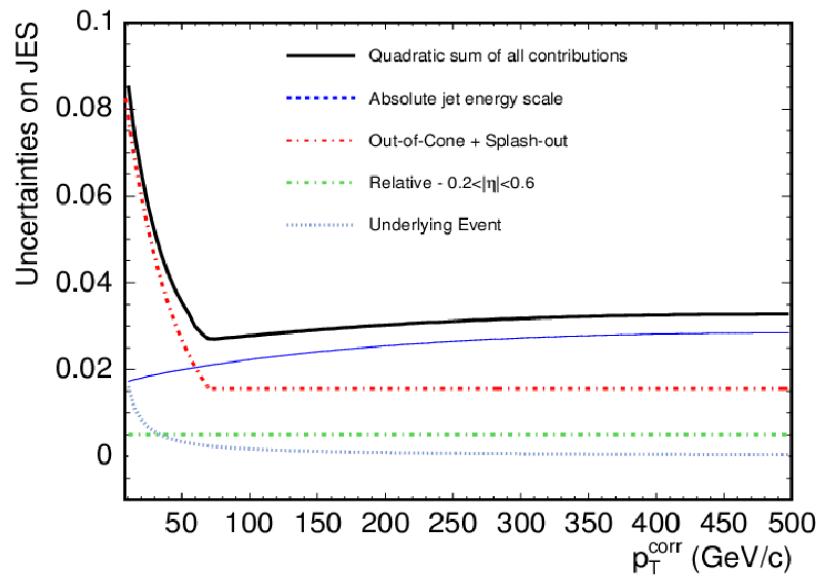
Transfer Functions and Efficiencies

- When we build the transfer functions, we need to take into account the fact that the TFs are not built from a perfect sample
- An additional jet-by-jet efficiency term is factored into the transfer functions to account for this. Efficiency is built in terms of parton P_T and proto-jet mass



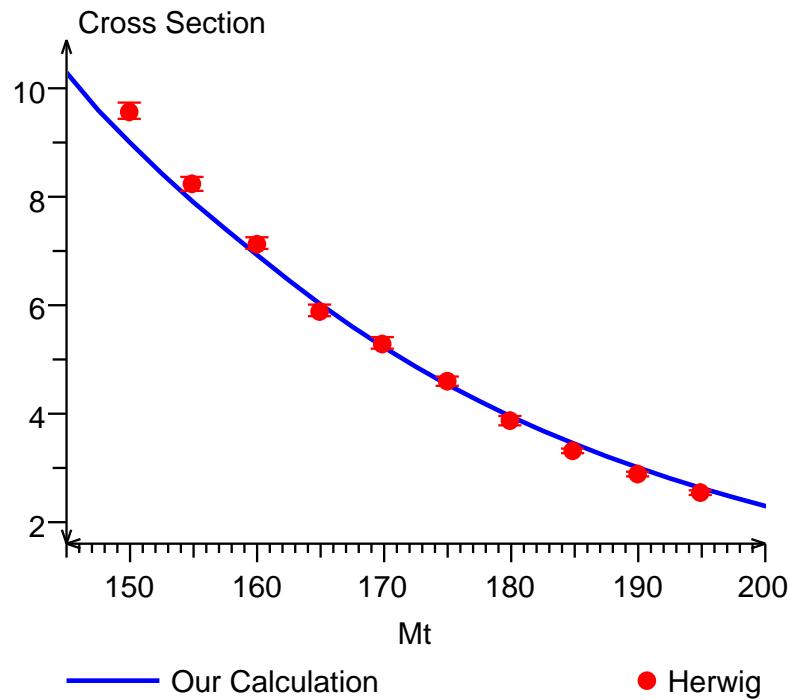
P_T -dependent JES vs. Constant JES

- Use of Δ_{JES} instead of a constant JES factor results in $\sim 15\%$ improvement in the expected stat + JES error
- It is not surprising that there is some change in the resolution: the transfer function Jacobian factor is different, so the likelihood is modified. But why such a big improvement?



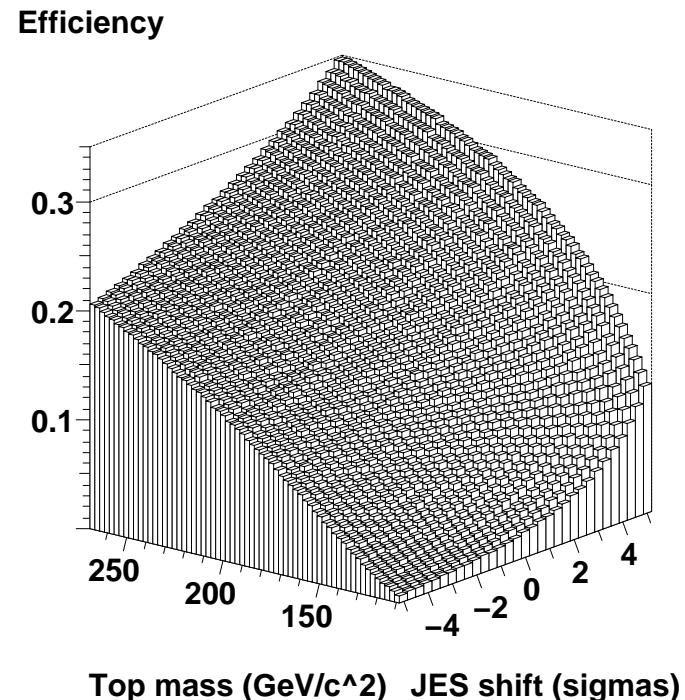
Normalization

- Our normalization is unchanged from MTM2.
- The normalization is obtained by integrating the matrix element and PDFs over the parton phase space. The blue curve shows the cross-section from our normalization compared to the Herwig cross-section (red dots).



Overall Acceptance

- We construct the acceptance the same way as in MTM2, except that the JES shift is performed in units of jet-dependent sigma rather than a flat factor.
- This results in a somewhat steeper slope with respect to JES.



Event Selection

- We use standard top selection rules plus exactly 4 tight jets and at least 1 b-tag. Only CEM, CMUP, and CMX are used — PHX electrons and BMU muons are removed.
- All samples are currently updated with latest versions and cuts:
 - JetUser jetCorr12
 - BTagObjects btag_1500invpb_v2
 - Good Run List version 18 (with bad beamline runs removed or recovered using the good run instructions)
 - Exactly 4 tight jets, where “tight” = 20 GeV at L5
 - MET > 20 GeV at L5
- LJ reclustering used except for QCD background, which uses DIL.
- Jet-based HF removal is performed on $W+HF/W+light$ backgrounds using methodII.cc code.

Data Samples

- We use all data up to period 12 (1.9 fb^{-1}).
- Note that the GRL v18 lists include instructions for recovering some events in the bad beamline runs, which also increases the number of events in periods 10 and 11.
- Total numbers agree with Tom's Method 2 calculations.

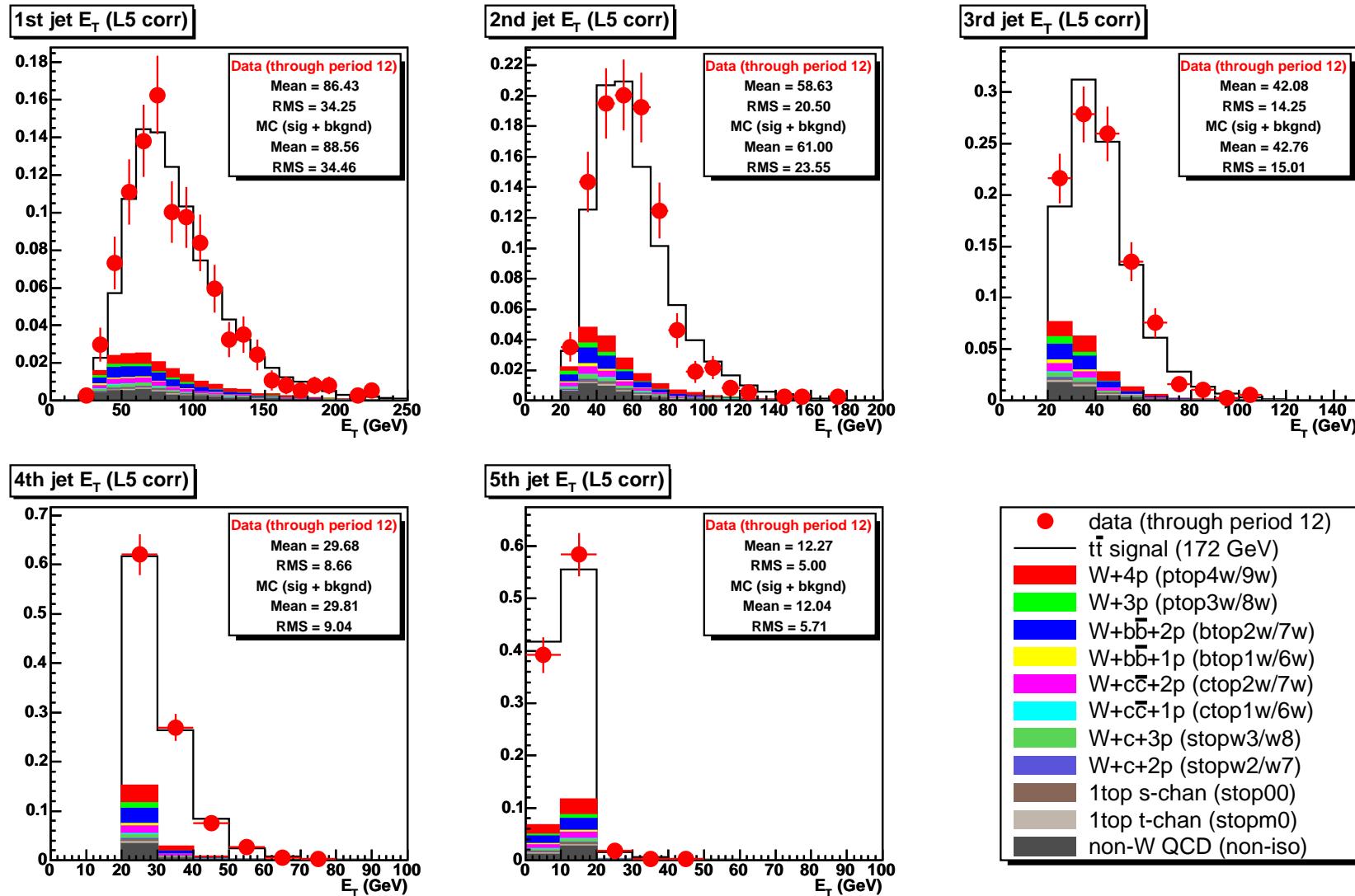
Period	Total tagged events	1 tag	≥ 2 tag
0d	70	56	14
0h	77	59	18
0i pre-shutdown	53	42	11
0i period 8	24	17	7
0i period 9	33	26	7
0i period 10	51	37	14
0j period 11	43	31	12
0j period 12	20	16	4
Total through period 12	371	284	87

Expected Background

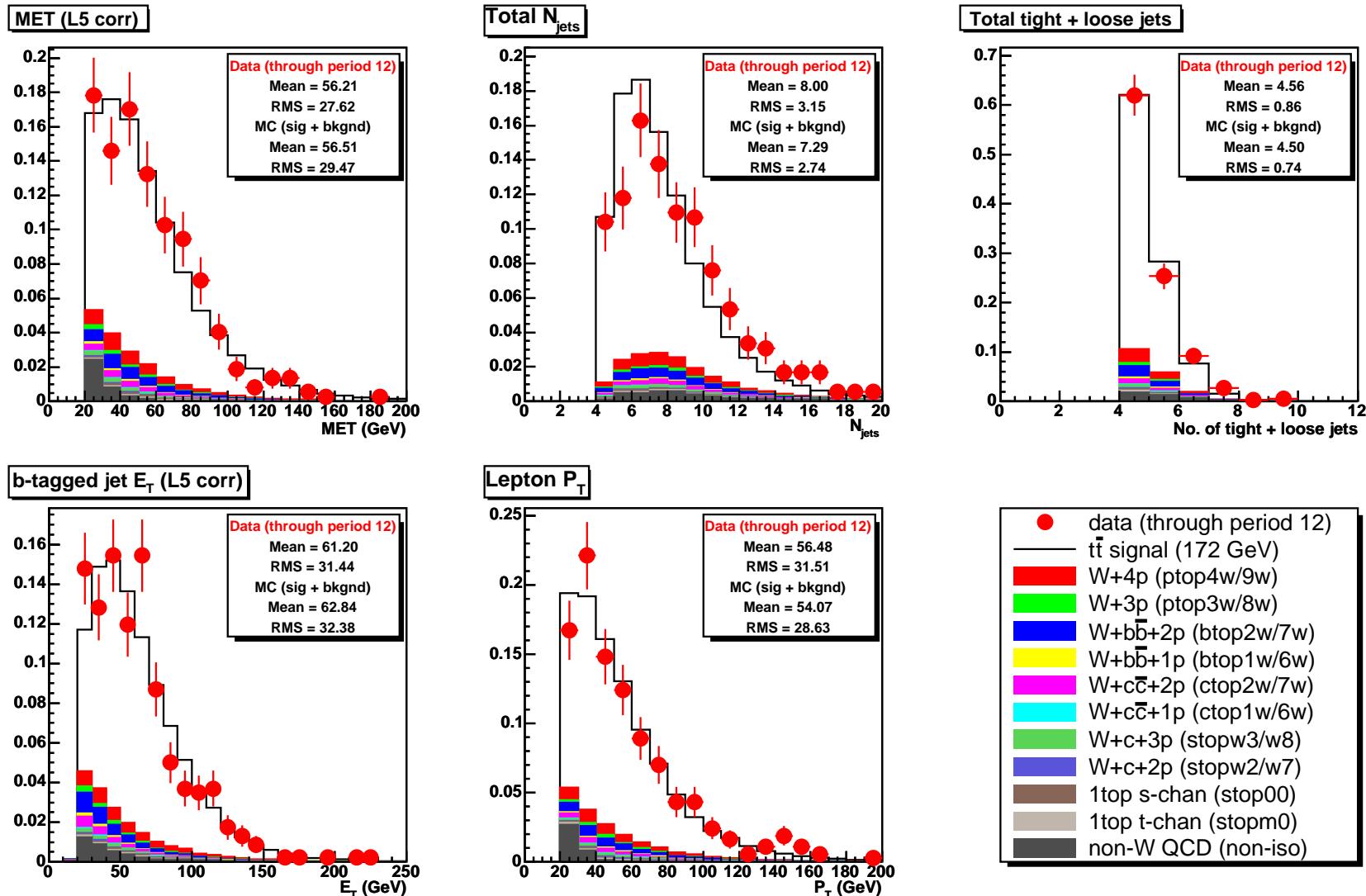
- We use numbers derived by Method II. The single-top samples are now used. The diboson and Z contributions are still absorbed into the W+light contribution.

Background	1 tag	≥ 2 tags
non-W QCD	13.81 ± 11.49	0.48 ± 1.50
W+light mistag	16.34 ± 3.56	0.31 ± 0.09
diboson (WW, WZ, ZZ)	3.29 ± 0.26	0.27 ± 0.03
$Z \rightarrow ee, \mu\mu, \tau\tau$	2.19 ± 0.26	0.19 ± 0.03
Sum of above 3	21.82 ± 3.58	0.77 ± 0.10
W+bb	13.75 ± 5.53	2.78 ± 1.14
W+c \bar{c} , c	12.31 ± 4.99	0.58 ± 0.24
Single top s-chan	1.49 ± 0.14	0.52 ± 0.07
Single top t-chan	1.53 ± 0.12	0.41 ± 0.05
Sum of above 4	29.08 ± 10.24	4.29 ± 1.36
Total background	64.71 ± 16.25	5.54 ± 2.56
Predicted top signal	182.63 ± 24.63	69.42 ± 11.19
Events observed	284	87

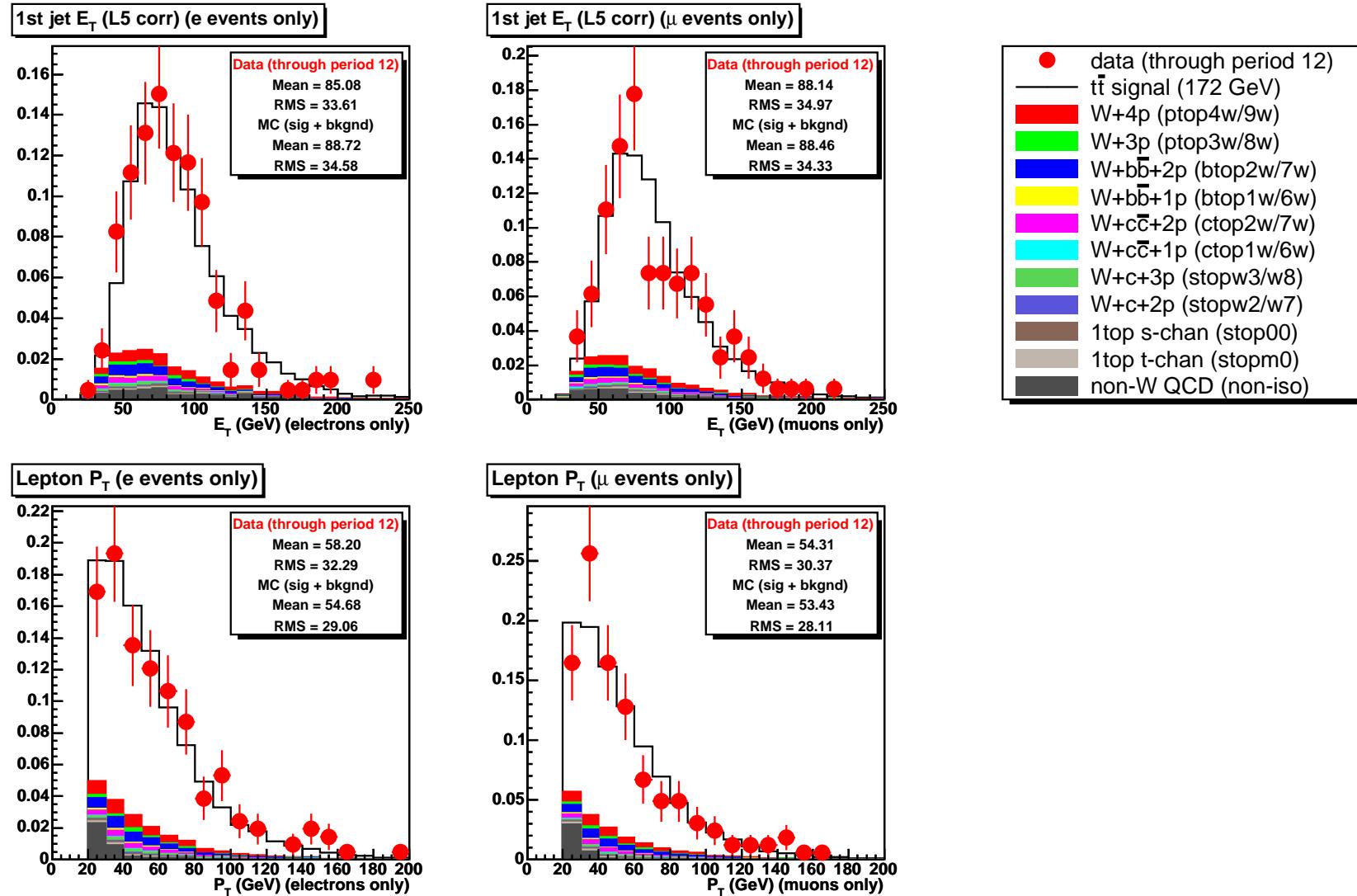
Data Validation (1)



Data Validation (2)



Data Validation (3)

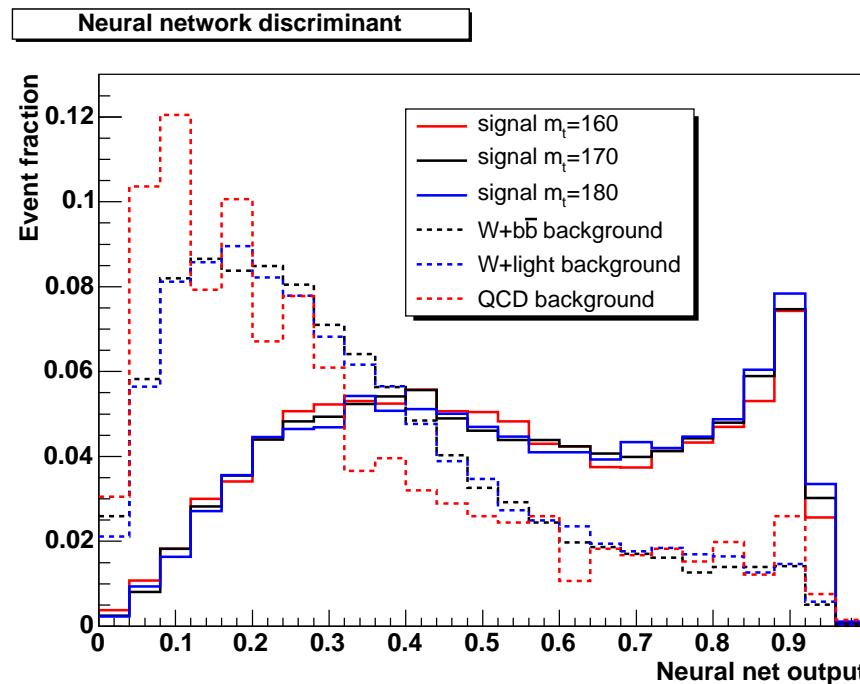


Left: electron events only, right: muon events only

Background Handling

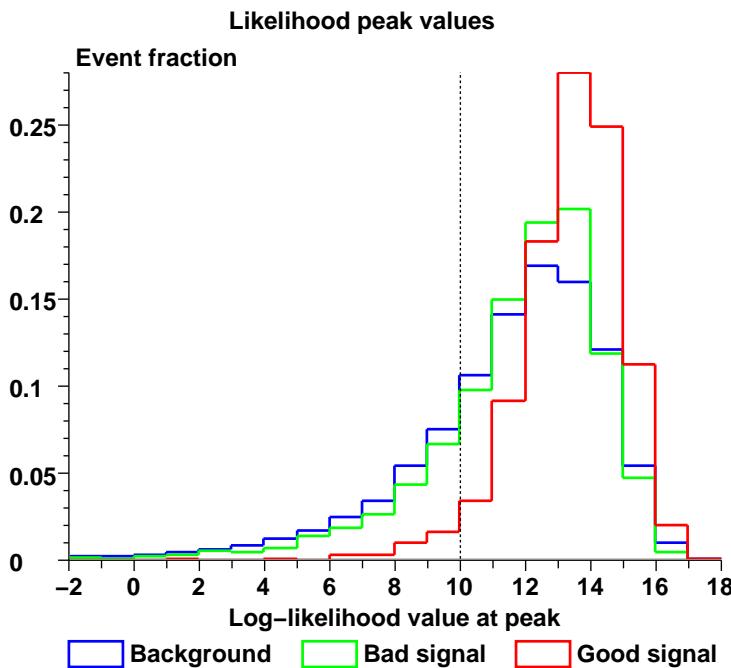
- Our background handling procedure has not changed. We subtract off the average expected background contribution (obtained from Monte Carlo) from the likelihood curve for a given event, where $f_{\text{bg}}(q)$ is the background fraction for an event obtained from the discriminant variable q .

$$\log L_{\text{sig}} = \sum_i \left[\log L_i - f_{\text{bg}}(q_i) \log \overline{L(\text{background})} \right]$$



Likelihood Cut

- We also employ a likelihood cut to deal with events which we call “bad signal” — events which are $t\bar{t}$ but where the 4 final jets are not the 4 jets from $t\bar{t}$ decay (extra jets from ISR/FSR, lost/merged jets, τ decay, misidentified dilepton/all-hadronic events, etc.) These make up $\sim 35\%$ of our signal!
- Since the likelihood has changed since MTM2, we adopt a new likelihood cut of 10 for MTM3. The expected resolution is improved by 12% w.r.t no-cut sample.



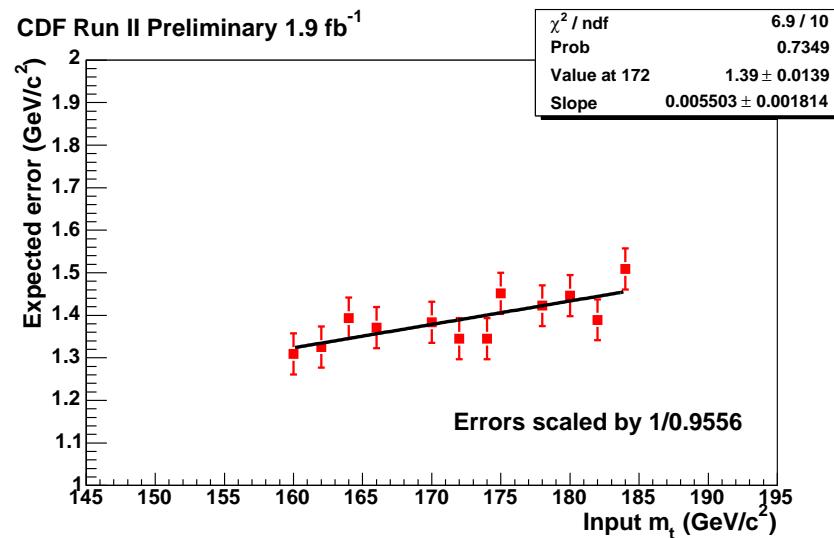
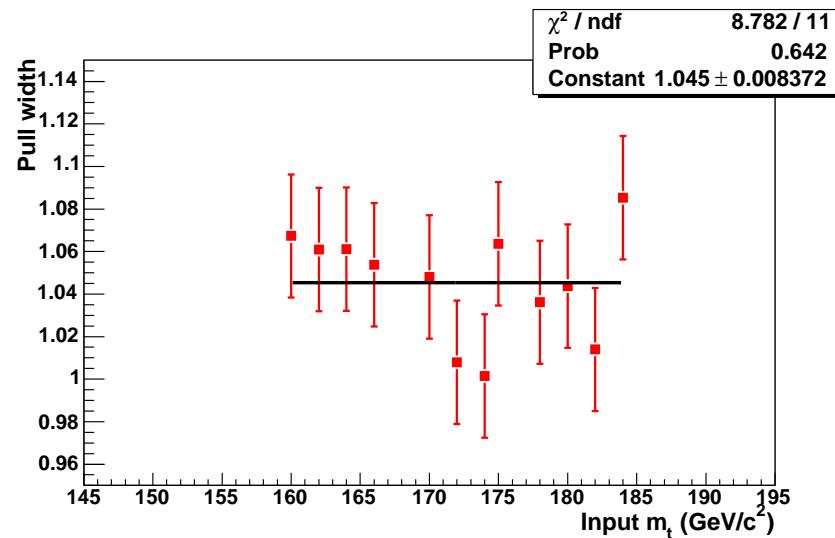
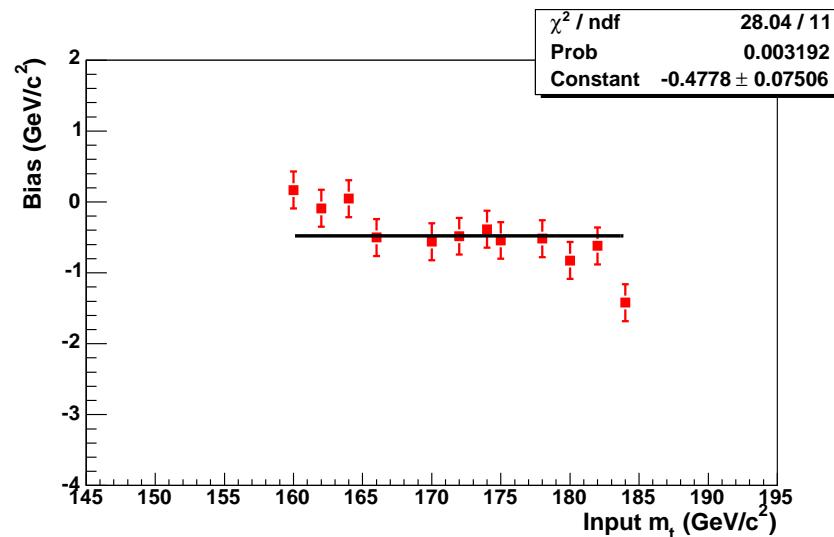
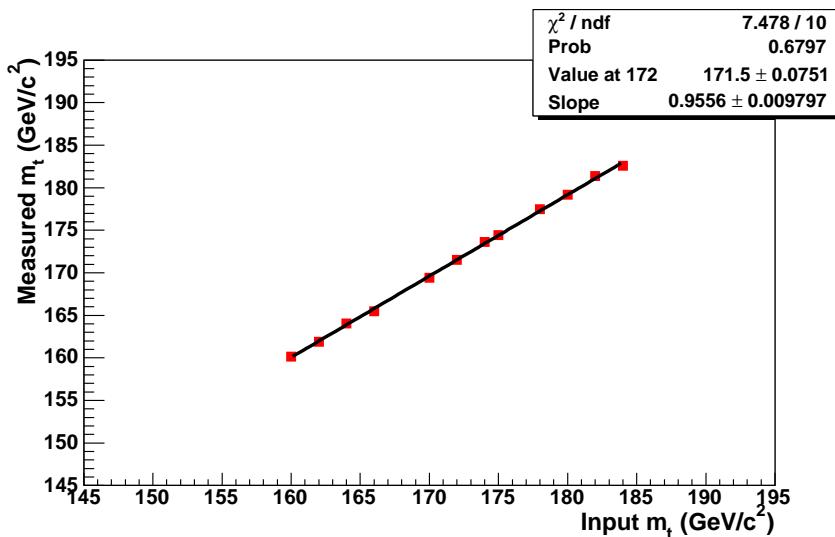
Pass rates:

Type of event	Total	1-tag	>1-tag
Good signal	96.6%	96.0%	98.0%
Bad signal	80.2%	80.5%	79.5%
Background	74.4%	74.5%	71.8%

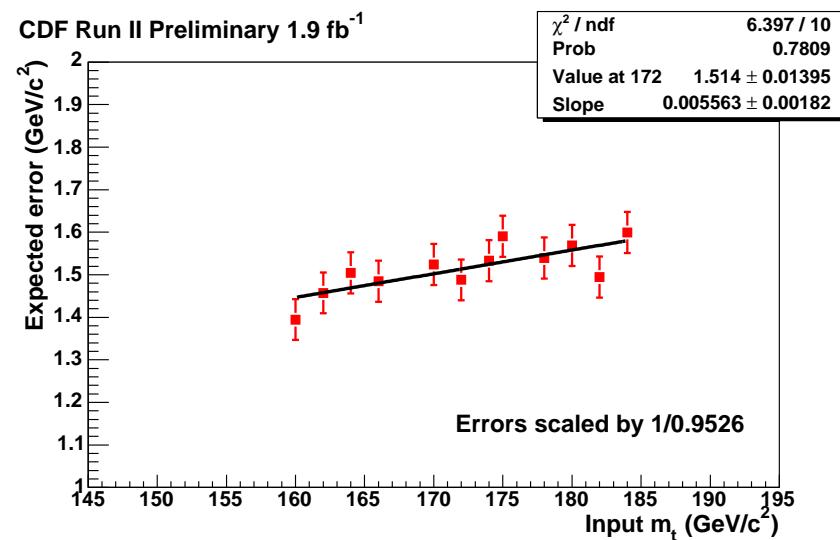
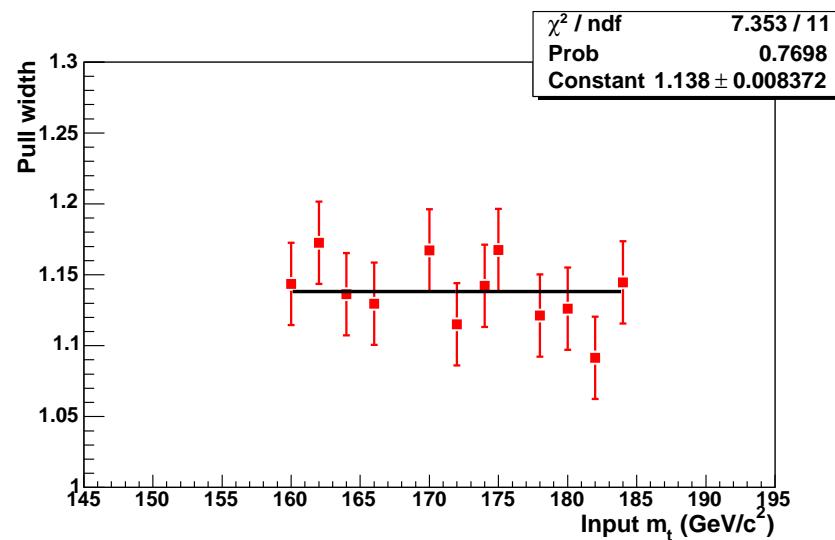
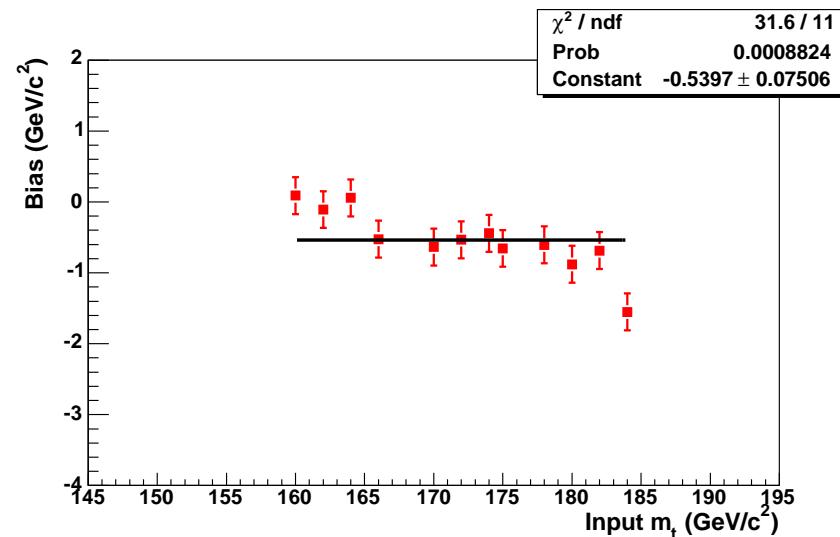
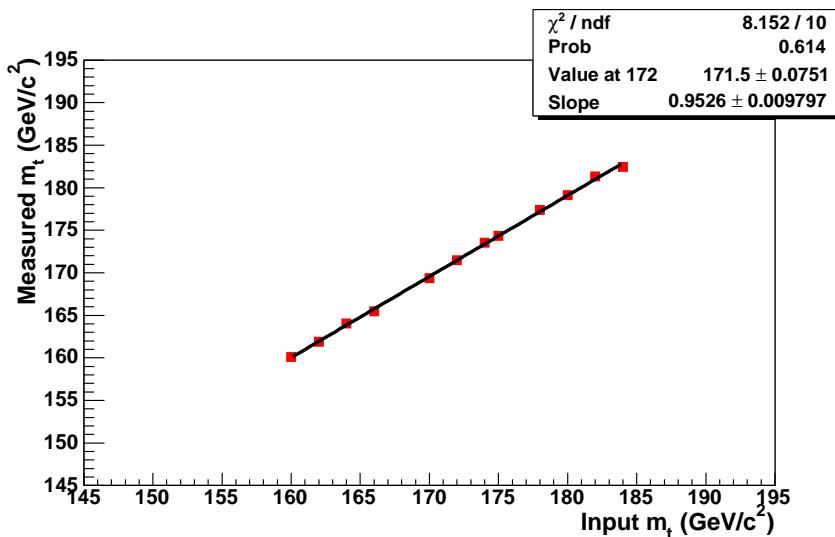
Pseudo-Experiments

- The PE procedure is largely the same as in MTM2.
- For each signal mass point, we run 2k PEs with 322.7 events/PE (expected number of events after likelihood cut is applied). Typically we integrate 8k events on the CAF for each mass point. Thus, a large amount of resampling takes place.
- The background contributes according to the Method II expected numbers described previously.
- Each individual sample's contribution is Poisson-fluctuated about the mean.
- PEs are run separately for 1-tag and >1-tag events (with the appropriate background fraction for each) and then combined. The profile likelihood is used to eliminate JES and obtain a likelihood in m_t .

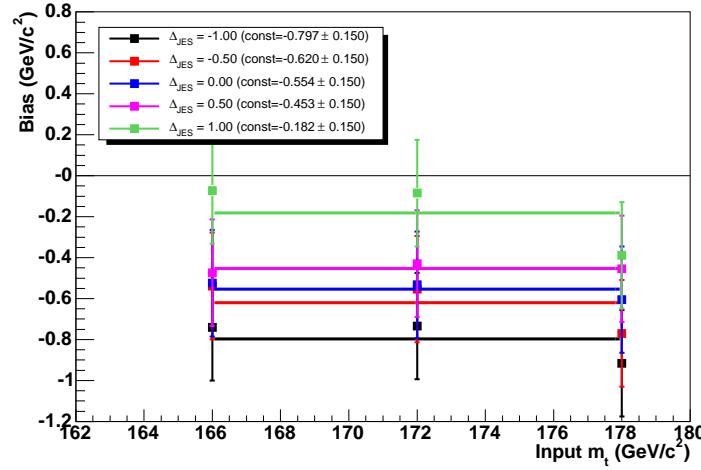
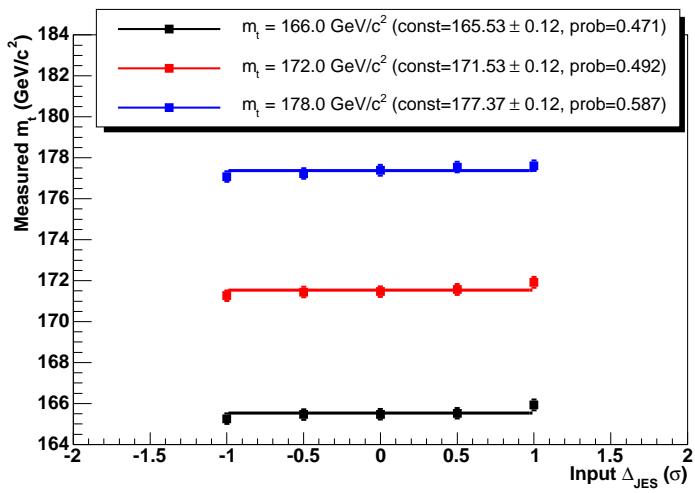
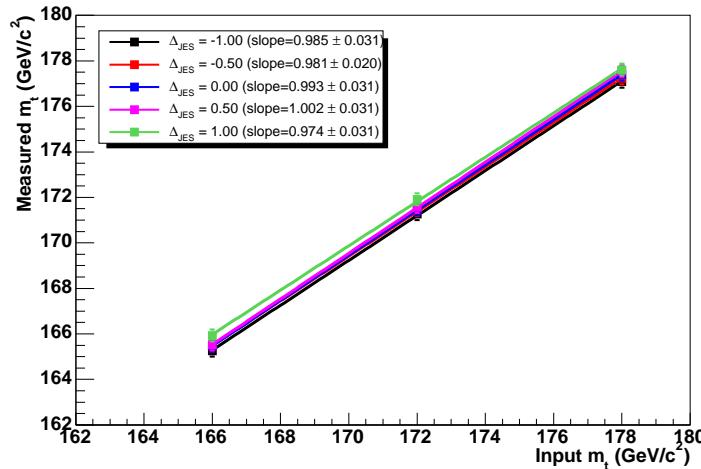
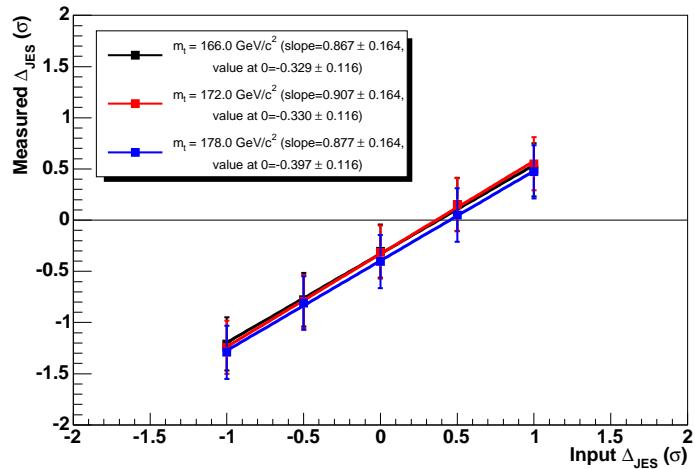
PE Results: Signal Only



PE Results: Signal + Background



JES Linearity Plots



Upper left: Output vs. input JES for different top masses. **Upper right:** Output vs. input mass for different JES. **Lower left:** Output mass vs. input JES for different masses. **Lower right:** JES bias vs. input mass.

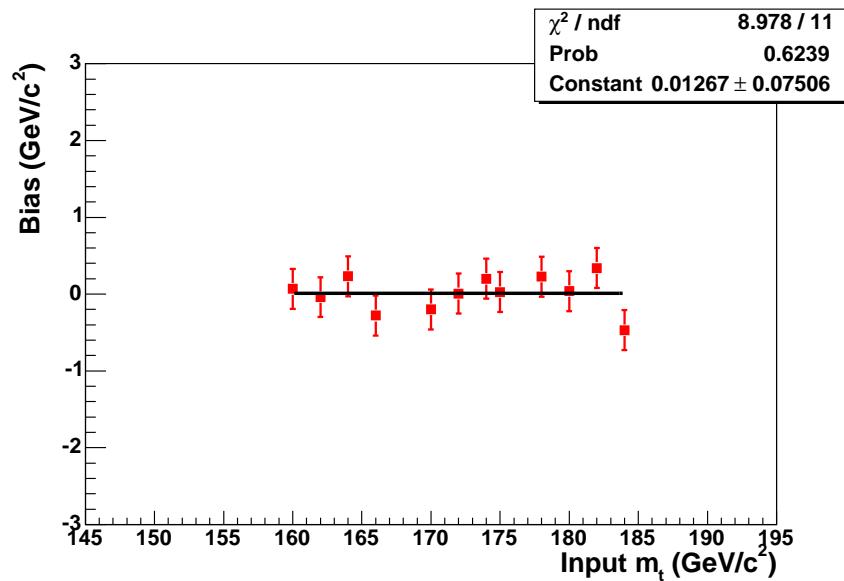
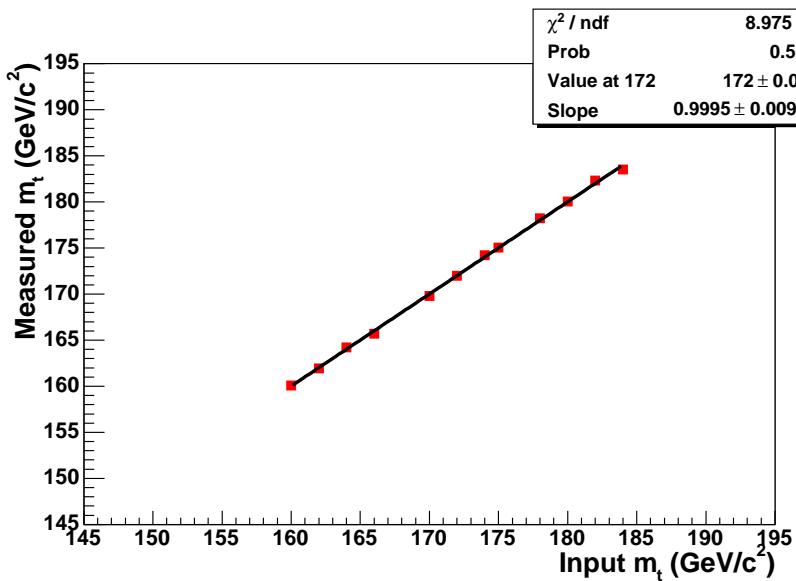
Calibration

- A bivariate linear mapping function is constructed to correct for the non-zero bias and the non-unit pull width of the maximum likelihood estimator:

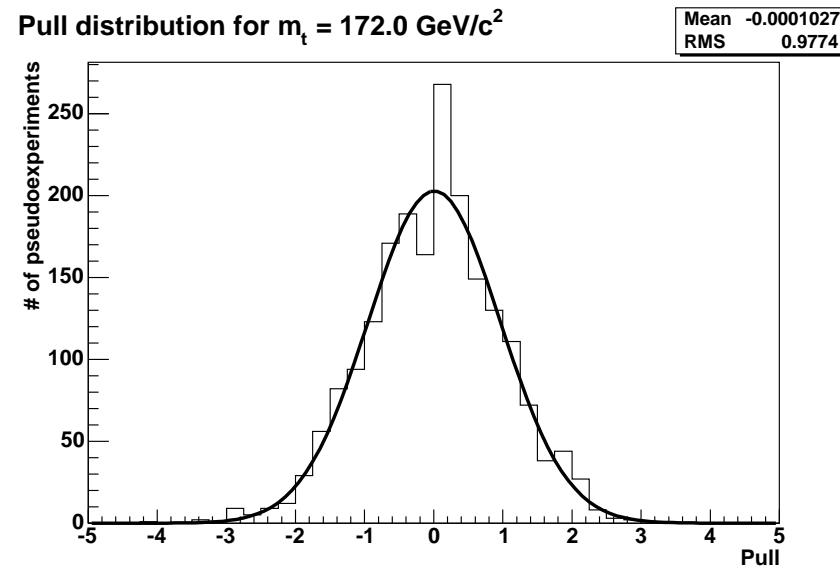
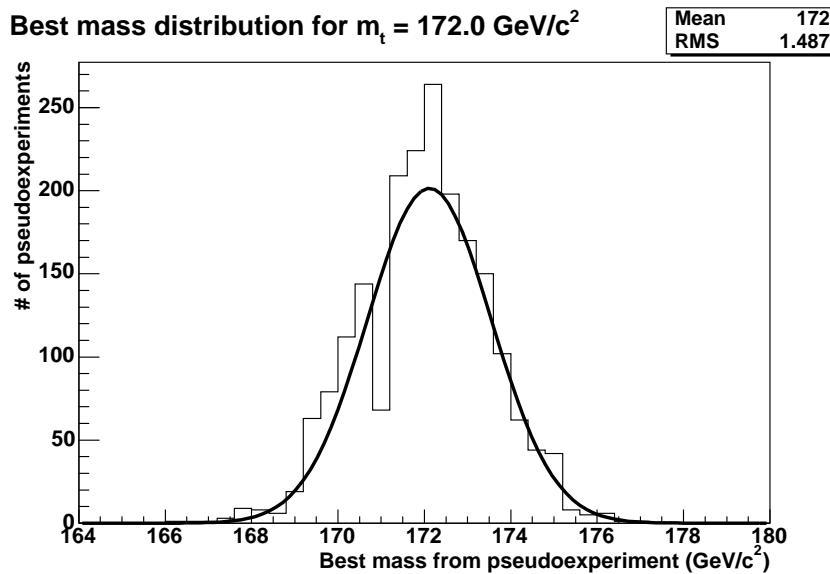
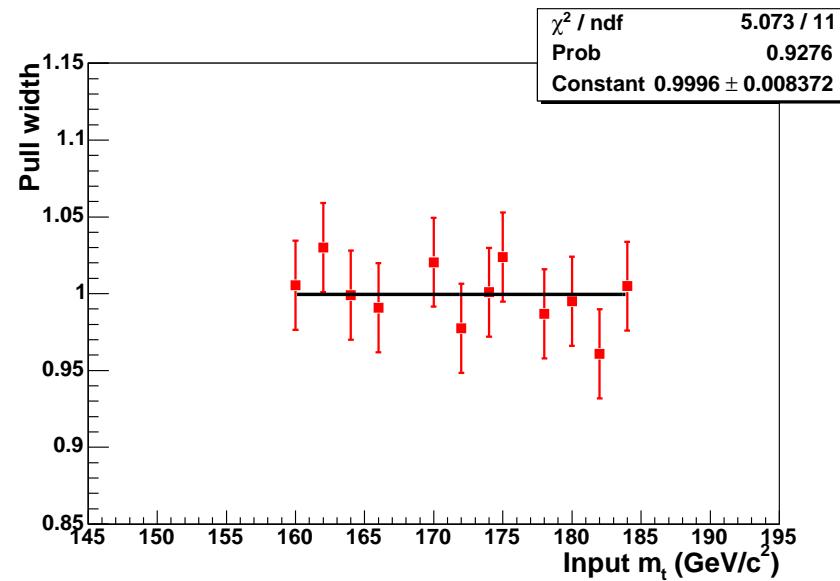
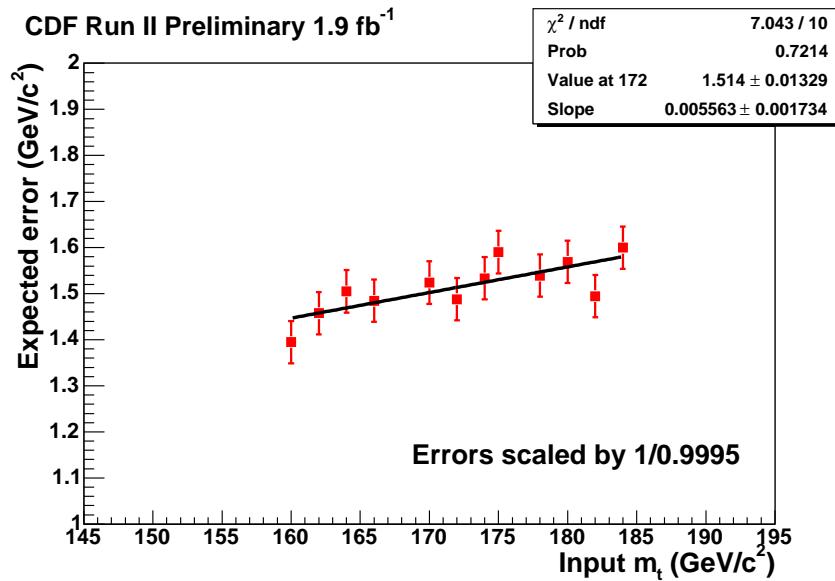
$$\Delta m_{estimated} = a \times \Delta m_{true} + b \times (\Delta_{JES})_{true} + c$$

$$(\Delta_{JES})_{estimated} = d \times (\Delta_{JES})_{true} + e \times \Delta m_{true} + f$$

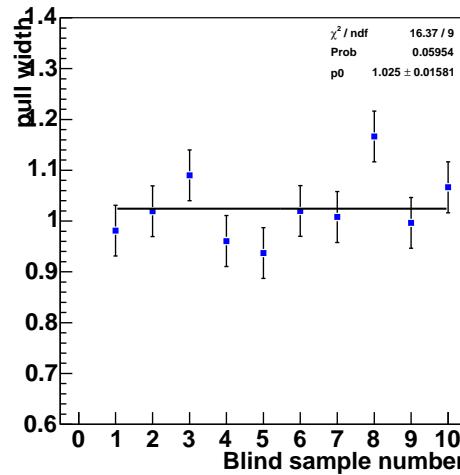
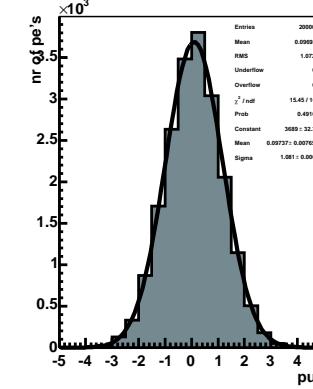
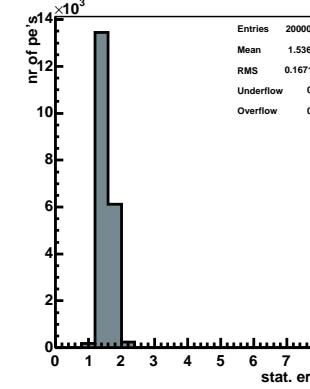
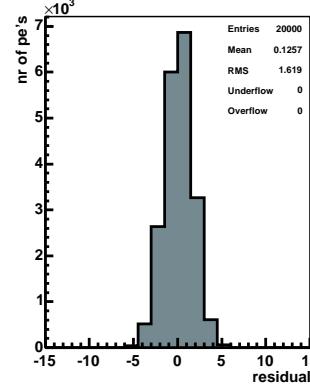
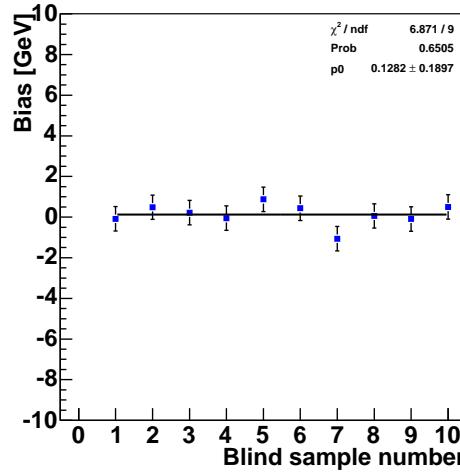
e is assumed to be 0. Other coefficients are fitted using the pseudo-experiment results, and the relationship is inverted to obtain mass and JES correction terms.



Calibrated Errors and Pulls



Blind Sample Results



- Top left: bias per blind sample; consistent with 0
- Bottom left: pulls per blind sample; consistent with 1
- Top right: distribution of residuals, errors, and pulls

Systematic Error Summary

Systematic source	Systematic uncertainty (GeV/c^2)
Calibration	0.13
MC generator	0.11 ± 0.37
ISR and FSR	0.50 ± 0.37
Residual JES	0.60
b-JES	0.36
Lepton P_T	0.18
Permutation weighting	0.01
Multiple interactions	0.05
PDFs	0.41
Background: fraction	0.27
Background: composition	0.24
Background: average shape	0.04
Background: Q^2	0.08 ± 0.07
Gluon fraction	0.00
Total	1.11

Systematic Errors: Residual JES

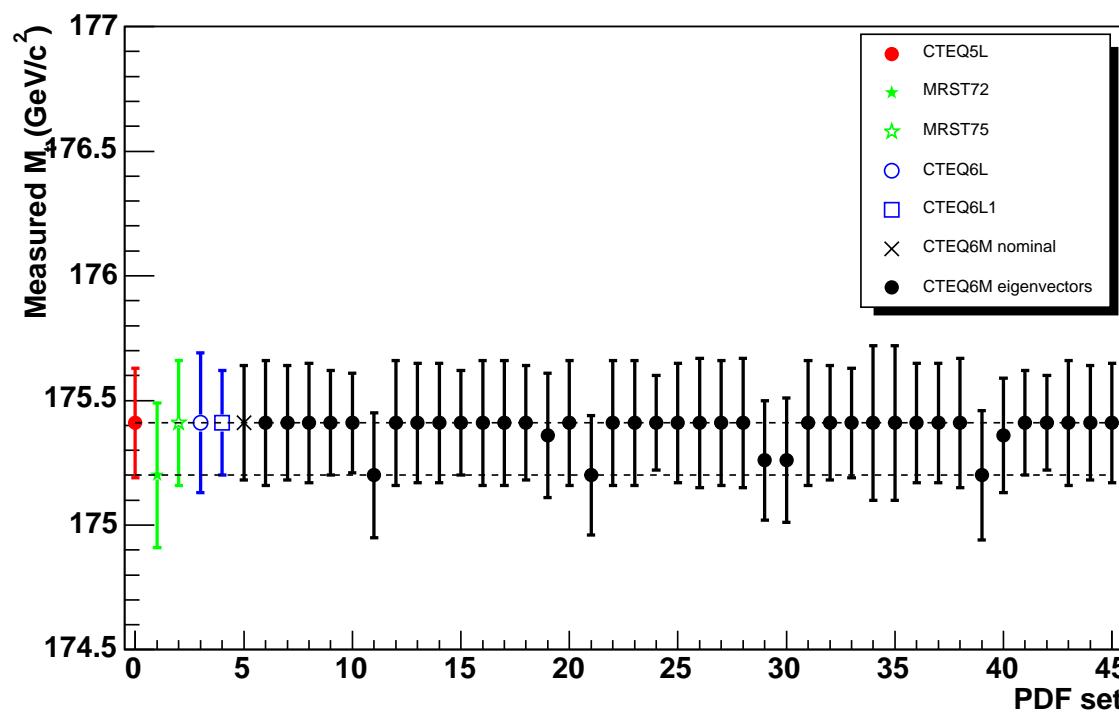
Systematics for the residual JES uncertainty. All samples are based on 5k events from the ttop72 sample with a nominal mass of $172 \text{ GeV}/c^2$.

Sample	Measured m_t (GeV/c^2)
Nominal	172.35 ± 0.37
Level 1 +1 σ	172.51 ± 0.38
Level 1 -1 σ	172.29 ± 0.37
Level 4 +1 σ	172.37 ± 0.39
Level 4 -1 σ	172.37 ± 0.37
Level 5 +1 σ	173.01 ± 0.39
Level 5 -1 σ	171.92 ± 0.36
Level 6 +1 σ	172.33 ± 0.38
Level 6 -1 σ	172.44 ± 0.37
Level 7 +1 σ	172.45 ± 0.37
Level 7 -1 σ	172.36 ± 0.39
Level 8 +1 σ	172.26 ± 0.41
Level 8 -1 σ	172.45 ± 0.37

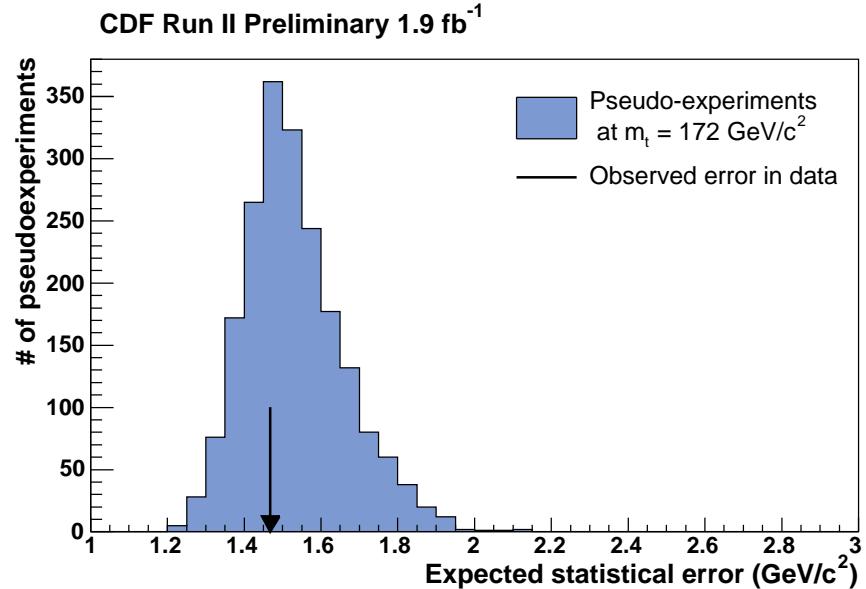
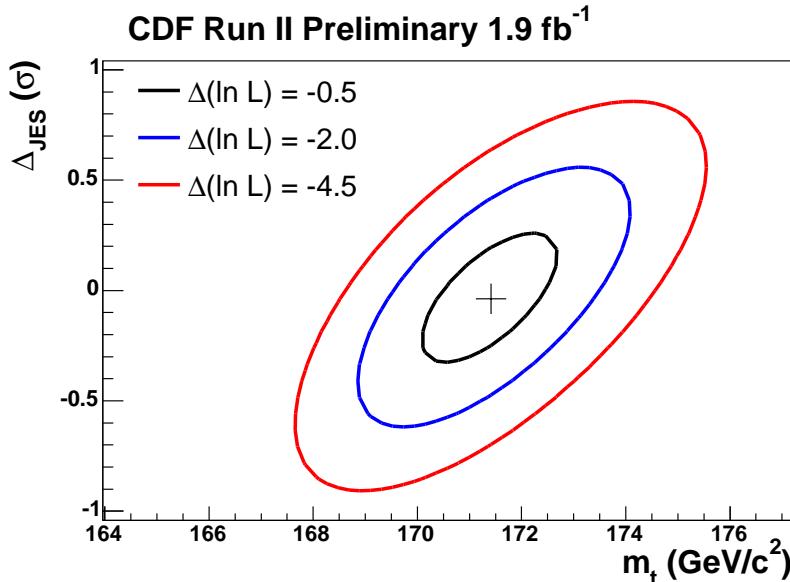
Systematic Errors: ISR/FSR and PDFs

ISR and FSR are varied together

Sample	Measured m_t (GeV/c^2)	Δm_t (GeV/c^2)
Nominal PYTHIA (ttop75)	175.12 ± 0.26	—
PYTHIA more ISR/FSR (otop03)	175.62 ± 0.26	0.50 ± 0.37
PYTHIA less ISR/FSR (otop04)	175.06 ± 0.26	-0.06 ± 0.37



Results with 1.9 fb^{-1}



$$m_t = 171.41 \pm 1.10 \text{ (stat)} \pm 0.98 \text{ (JES)} \pm 1.11 \text{ (sys)} \text{ GeV}/c^2$$

$$m_t = 171.41 \pm 1.84 \text{ (total)} \text{ GeV}/c^2$$

$$\Delta_{\text{JES}} = 0.03 \pm 0.31$$

- A lot more details about this analysis can be found in the CDF note 9245 and on the analysis web page at

http://www-cdf.fnal.gov/internal/physics/top/run2mass/multivar_analysis/mtm3/